



Statistiques pour biologistes

Atelier de formation – 25 mai 2017

Journée de la recherche du CHU de Québec – Université Laval

Anne-Sophie Julien

Plateforme de recherche clinique

anne-sophie.julien@crchudequebec.ulaval.ca



Plateforme de recherche clinique

Services offerts

- Méthodologie et biostatistique (58\$/heure)
 - Conception et préparation du projet de recherche (question de recherche, choix des paramètres, devis, rédaction protocole, questionnaire)
 - Gestion des données (grille de collecte, codification, saisie, validation)
 - Biostatistiques (protocole, design expérimental, calcul de taille d'échantillon, randomisation, analyses statistiques, interprétation, rédaction & révision article)
- Ateliers de formation (gratuits pour les membres du CRCHU)
 - 01 Conception projet de recherche, 02 analyses statistiques de base, 03 calcul de taille d'échantillon, 04 création d'une banque de données Excel, 05 étapes administratives, 06 BPC & MON, Budget de recherche, 10 développement de questionnaire et instrument, 11 études diagnostiques, 07 score de propension, 09 validation de modèles paramétriques, 08+12à15 introduction à SPSS, R et SAS, analyses statistiques avec SPSS, R et SAS



Atelier 09 – Validation de modèles paramétriques

– Suite de l’atelier d’aujourd’hui

- Date: 8 juin 14h-16h
- Lieu: CHUL, local H00-329
- Contenu
 - Résidus
 - Postulats fréquents
 - Validation ANOVA, régression linéaire & régression logistique

- Inscription sur:

<http://rce.crchudequebec.ulaval.ca/>



Question 1

Je ne sais pas si je dois faire un test unidirectionnel ou bidirectionnel.



Uni vs Bi latéral

Exemples: comparaison de moyennes et test de corrélation

- Test unidirectionnel -> unilatéral (1 tail)
 - H0: Moyenne \leq Moyenne2 / Corrélation ≤ 0.30
 - H1: Moyenne1 $>$ Moyenne2 / Corrélation > 0.30
- Test bidirectionnel -> bilatéral (2 tails)
 - H0: Moyenne 1 = Moyenne2 / Corrélation = 0
 - H1: Moyenne 1 \neq Moyenne2 / Corrélation $\neq 0$



Question 2

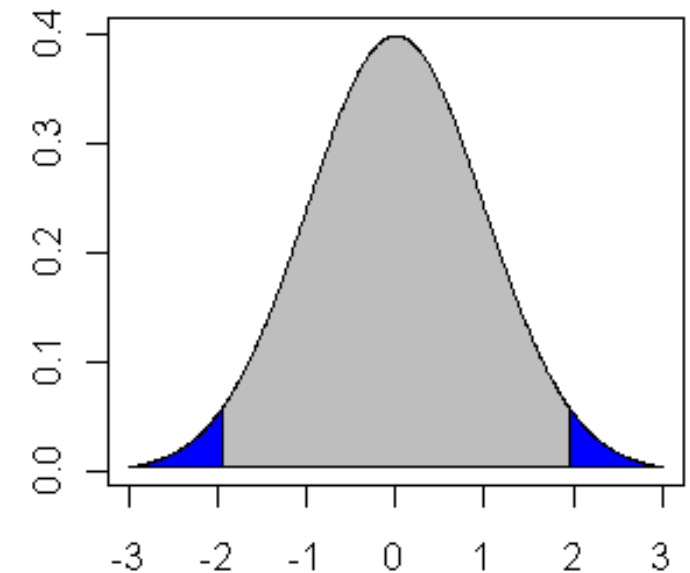
$$P > 0.05$$

Puis-je conclure que les deux groupes sont égaux?



Tests de supériorité et d'égalité

- Test de supériorité (ou différence)
 - H_0 : Moyenne1 = Moyenne2
 - H_1 : Moyenne1 \neq Moyenne2
 - > On rejette ou non H_0 , mais on ne l'accepte pas.
- Test d'égalité
 - H_0 : Moyenne1 \neq Moyenne2
 - H_1 : Moyenne1 = Moyenne2
 - > En rejetant H_0 , on peut conclure à l'égalité.
- Test de non-infériorité





Question 3

*Quelle est la différence entre SD et SEM?
Et lequel est le mieux pour publier?*



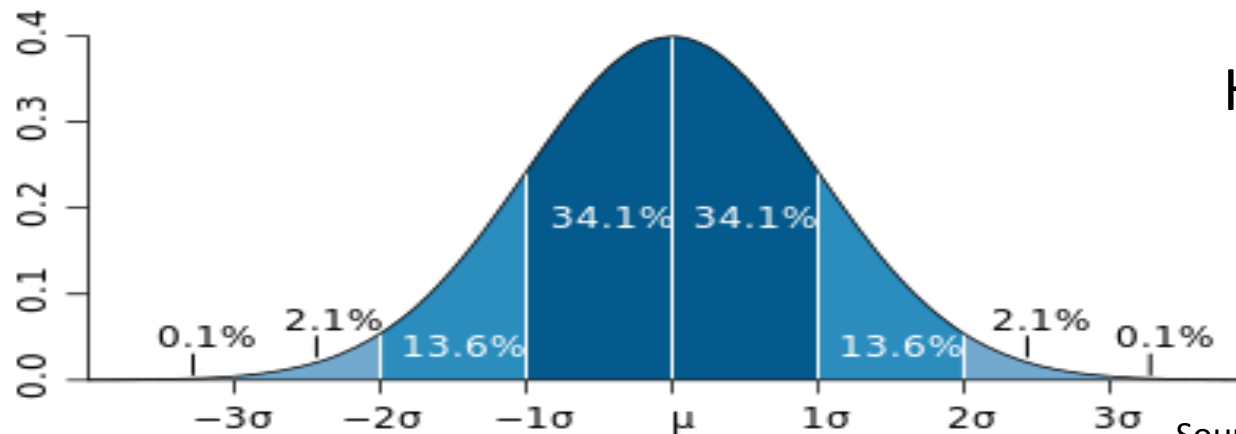
Statistiques descriptives pour variables continues

Mesure de localisation

- Moyenne
- Médiane

Mesures de dispersion associées à la moyenne

- Variance : une mesure de dispersion des données autour de la moyenne
- Écart-type : racine carré de la variance = σ



Histogramme d'une
loi normale



Mesures de dispersion

Écart-type (Standard Deviation)

- Indicateur de la variabilité des données
- Présenté dans les statistiques descriptives

Erreur-Type (Standard Error of the mean)

- Écart-type divisé par la racine carrée de N
- Mesure l'incertitude associée à l'estimation de la moyenne
- Présentée dans les résultats d'analyses

Référence: Nagele, P. (2001). Misuse of standard error of the mean (SEM) when reporting variability of a sample. A critical evaluation of four anaesthesia journals. *British Journal of Anaesthesia*, 90 (4): 514-16.



Question 4

Comment définir le n?

- *J'ai un puit, mais un qPCR en triplicata.*
- *J'ai une souris, mais 2 PCR*



Types d'unités

Unité expérimentale

- Le traitement y est appliqué
- Indépendantes

Unité échantillonnale

- Les mesures y sont prises
- Dépendantes
- Peut être la même que l'unité expérimentale, ou différente.

- Exemple 1: Le plasma du patient 001 est stimulé par le produit X. On mesure la concentration d'une certaine protéine. L'unité échantillonnale est la même que l'unité expérimentale, le plasma du patient 001.
- Exemple 2: Le plasma du patient 001 est stimulé par le produit X, puis distribué dans 3 puits. On mesure la concentration d'une certaine protéine dans chaque puit. L'unité échantillonnale est le puit. Il y a une corrélation entre les unités échantillonnales d'une même unité expérimentale (plasma 001).



Répétitions

- Lorsqu'un même traitement est appliqué à plusieurs unités expérimentales, alors ces dernières sont considérées comme des répétitions du traitement.
- Chaque répétition est indépendante
- Plusieurs répliques font qu'une même combinaison de variables peut être présente pour plusieurs individus.
- Permet d'estimer l'erreur expérimentale. Sans réplique, on ne peut faire de tests (sauf pour certains dispositifs particuliers)



Pseudos-répétitions

- Lorsque l'unité expérimentale n'est pas la même que l'unité échantillonnale, il s'agit de pseudos-répétitions. Si plusieurs unités échantillonnales de même nature (par exemple des puits avec plasma) se retrouvent dans une même unité expérimentale soumise à un traitement unique, alors ces unités sont des pseudos-répétitions
- Chaque unité échantillonnale n'est pas indépendante.
- Modélisation
 1. Calculer la moyenne des pseudos-répétitions pour n'avoir qu'une observation par unité expérimentale -> données indépendantes obtenues -> modèles statistiques plus simples, mais légère perte d'information
 2. Utiliser toutes les pseudos-répétitions, mais en tenant compte de leur dépendance dans le modèle. -> pas de perte d'information, mais modèles statistiques plus complexes



Répétitions vs pseudo-répétitions

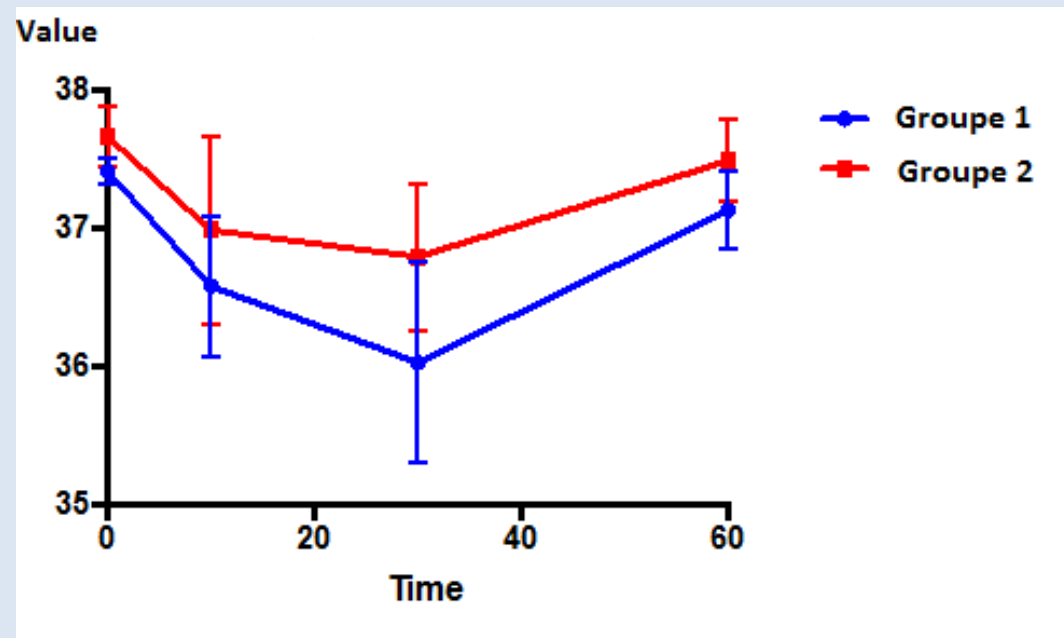
Habituellement, il y a une plus grande variabilité entre les vraies répétitions que les pseudo-répétitions. Dans le cas contraire, le nombre d'unités échantillonnales peut être augmenté pour mieux modéliser cette variabilité.

Autrement, si on doit augmenter la taille d'échantillon, on augmente le nombre de répétitions (unités expérimentales).



Question 5

J'ai des données dans un graphique. Comment choisir quel test faire pour obtenir une valeur p?





Choix d'analyse

Plusieurs facteurs à considérer:

- L'objectif
- La nature de la variable dépendante (ce qu'on tente d'expliquer)
- La nature de la variable indépendante (les facteurs prédictifs)
- Le design expérimental
- Le type de données
 - Non corrélées (répétitions seulement)
 - Corrélées (présence de pseudos-répétitions ou données pairées, appariées)
- Le nombre de variables
 - Univariée (1 variable: distribution, descriptif)
 - Bivariée (2 variables: association)
 - Multivariée (>2 variables: association): plusieurs facteurs, sélection, confusion



Pour comparer des groupes (moyennes) avec une VD continue
 Pour voir l'effet de facteurs sur une VD continue

Tests	V Dépendante	V indépendante(s)	Type de données	Type d'analyse
Test T	Continue	Dichotomique	Non corrélées	Bivariée
Test T Pairé	Continue	Dichotomique	Corrélées	Bivariée
ANOVA (à 1 facteur)	Continue	Qualitative (inter)	Non corrélées	Bivariée
ANOVA (à 2 facteurs)	Continue	2 qualitatives(inter)	Non corrélées	Multivarié
ANOVA à mesures répétées (à 1 facteur)	Continue	Qualitative (Facteur intra)	Corrélées	Bivariée
ANOVA à mesures répétées (à 2 facteurs)	Continue	2 qualitatives (1 inter + 1 intra ou 2 intra)	Corrélées	Multivarié



Pour comparer des groupes (distributions) avec une VD continue

Pour voir l'effet d'un facteur sur une VD continue

-> Normalité non respectée

Tests	Variable dépendante	Variable indépendante	Type de données	Type d'analyse
Wilcoxon	Continue	Dichotomique	Non corrélées	Bivariée
Wilcoxon Rangs signés	Continue	Dichotomique	Corrélées	Bivariée
Kruskall-Wallis	Continue	Qualitative (1 facteur inter)	Non corrélées	Bivariée
Friedman	Continue	Qualitative (1 facteur intra)	Corrélées	Bivariée



Question 6

Je veux comparer la moyenne de deux groupes, mais j'ai un petit échantillon. Est-ce que je peux utiliser le test de Wilcoxon?

Les postulats du test T ne sont pas respectés, donc je peux faire un test de Wilcoxon?



Test de Wilcoxon

- Wilcoxon n'est pas toujours la solution
- Objectifs distincts
 - 2 distributions avec même forme, homogénéité des variances -> test d'égalité des médianes
 - 2 distributions avec différentes formes, hétéroscédasticité -> test de dominance stochastique
- Variables continues
- Valable aussi pour Kruskal-Wallis ($k > 2$)
- Petit échantillon, Wilcoxon pas mieux



Question 7

Comment choisir entre l'ANOVA à 1 facteur (1-way), l'ANOVA à 2 facteurs (2-ways) ou les comparaisons multiples?

On m'a déjà dit que lorsque le P de l'ANOVA était significatif, on pouvait ensuite procéder sans contrainte à des test de student par paires de groupes. Est-ce vrai ?

Comment choisir le post-test d'une ANOVA?



Réponse 7

ANOVA à 1 facteur

- VD = continue, VI = 1 facteur, variable qualitative (groupes)
- Objectif: Différence de moyennes entre les groupes?

ANOVA à 2 facteurs

- VD = continue, VIs = 2 facteurs, variables qualitatives (groupes)
- Objectifs: Différence de moyennes entre les groupes? Y-a-t-il un effet du facteur 1? Du facteur 2? Une interaction entre les 2?

Comparaisons multiples

- Si un test global d'ANOVA est significatif (ie il existe une différence de moyennes), où se trouve la différence
- Plusieurs tests effectués, augmente le risque d'erreur de Type I
- Ajustement/correction à effectuer



ANOVA

Exemple

On mesure le VO₂Max de plusieurs patients, puis on les divise en trois groupes selon leur IMC.

On désire savoir si le VO₂Max est affecté par la tranche d'IMC du patient.

Source	DDL	Type III SS	Moyenne quadratique	Valeur F	Pr > F
imc	2	158.7053333	79.3526667	11.54	0.0016

P = 0,0016: Le VO₂Max diffère pour au moins un des trois groupes au seuil de 5%.

Le groupe "Surpoids" a un VO₂Max significativement inférieur aux deux autres.

imc	vo2max LSMEAN	Nombre LSMEAN
Insuffis	43.5000000	1
Normal	46.3000000	2
Surpoids	38.4400000	3

Least Squares Means for effect imc
Pr > |t| for H0: LSMean(i)=LSMean(j)
Dependent Variable: vo2max

i/j	1	2	3
1		0.1171	0.0101
2	0.1171		0.0005
3	0.0101	0.0005	



Comparaison multiples

- Si test global d'ANOVA SIG -> comparaisons multiples par paire pour identifier entre quels groupes il y a des différences significatives
- Corrections / ajustement: le seuil combiné doit rester près de 5%
- Alpha combiné = $1 - (1 - \alpha_{\text{par comparaison}})^k$, avec k comparaisons

Quelques choix d'ajustement

- Exploratoire; aucun ajustement: Fisher's LSD
- Traitements vs contrôle: Dunnett
- Données balancées: Tukey
- Faible nombre, planifiées: Bonferroni
- Snooping, grand nombre, non planifiées: Scheffe
- Génétique, très grand nombre: Benjamini and Hochberg
- ... beaucoup d'autres existent

Référence:

Milliken, G.A. and D.E. Johnson (2009).
Analysis of Messy Data Volume 1:
Designed Experiments, 2nd edition.
Chapman and Hall/CRC. 674 pages.



Question 8

Un test ANOVA dans Prism me confirme que les moyennes entre les groupes sont différentes (Are means signif. different ? $P < 0.05$). Par contre les post-test Tukey ou Bonferroni ne me donne pas de différence significative entre tous les groupes.

Le test Newman-Keuls y parvient cependant. De même, si je fais un test de student entre les deux groupes problématiques en ANOVA, la différence est significative.

Dans mes données la variance est très grande (la déviation standard est plus grande que la moyenne).

Que faire dans une telle situation ?



Réponse 8

- Possibilité de variables influentes si variance énorme
- Postulats de l'ANOVA respectés (normalité & hétérogénéité)?
- Test global: est-ce qu'au moins l'une des moyennes est différente des autres?
- Si test global NS, pas de comparaisons multiples
- Si test global SIG, comparaisons multiples
 - Sans correction = test T, probablement une significative
 - Avec correction, pourraient toutes être NS si test global près de 5%
- Bonferroni, particulièrement conservateur si # comparaisons élevé
- Newman-Keuls, similaire à Tukey, mais moins conservateur (donc + de résultats SIG)



Question 9

Outre augmenter le nombre d'échantillons analysés y-a-t'il d'autres moyens de parvenir à des différences significatives, exemple faire les analyses statistiques sur les données transformées en logarithme ? Si oui, comment justifie t'on ce type de transformation ?



Réponse 9

- Puissance \uparrow si $N \uparrow$, variance \downarrow ou effet / différence \uparrow
- Modèle correctement spécifié, test plus puissant
- Validation de postulats

- Transformation de Box-Cox pour Y
 - Exemple: concentrations \rightarrow log
 - Aussi: racine carrée, inverse, au carré, etc.
 - Utile avec problèmes de normalité des résidus
- Interprétation commune



Validation de modèles paramétriques

– Postulats

- Hypothèses du modèle doivent être respectés
- Dépendent du modèle
- Normalité des résidus
- Homoscédasticité des résidus (variances égales)
- Linéarité
- Indépendance (absence d'auto-corrélation)
- ...

– Sources de biais

- Valeurs influentes
- Multicolinéarité
- Données manquantes

– Pouvoir prédictif du modèle

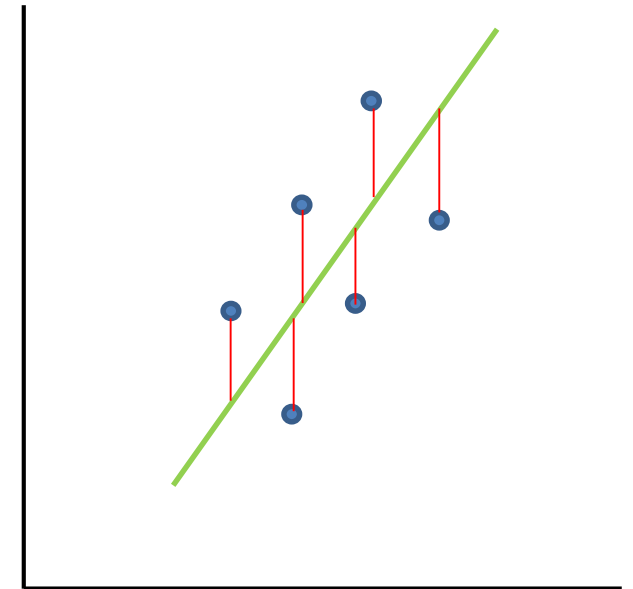
- Pouvoir explicatif des variables explicatives
- Validation des prédictions
- Validation croisée

Sans validation, la conclusion d'une analyse pourrait être erronée!



Résidus

- Résidu ordinaire: $e_i = Y_i - \hat{Y}_i$
 - Où Y_i = valeur observée et \hat{Y}_i = valeur prédite
 - Bon modèle: Résidus tous près de 0
- Résidu studentisé: e_i divisé par son écart-type
 - Ramène les résidus sur une échelle connue: scores Z





Normalité des résidus

- Les modèles supposent que les résidus suivent une distribution normale de moyenne 0 et de variance σ^2 (homoscédasticité)
- Analyses robustes à une légère déviation de la normalité
 - Résultats extrêmes moins affectés
 - Grand kurtosis → conservateur
- Test T, ANOVA, Régression linéaire, Modèle linéaire mixte
- Validation graphique, par des statistiques ou des tests



Homogénéité des variances

- Les modèles supposent que les résidus suivent une distribution normale de moyenne 0 et de variance σ^2 (homoscédasticité)
- Lorsque la variance des résidus diffère selon les groupes ou la valeur de Y, il y a hétéroscédasticité
- Test T, ANOVA, Régression linéaire & logistique
- Validation graphique ou par des tests



Linéarité

- Le modèle de régression linéaire suppose qu'il existe une relation linéaire entre VD et VI

$$VD = \beta_0 + \beta_1 VI$$

- Le modèle de régression logistique suppose une relation linéaire entre le logit de la probabilité de succès (π) et la VI

$$\log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 VI$$

- Si la relation n'est pas linéaire, le modèle ne sera pas un bon indicateur de la relation
- Vérification graphique ou par test d'ajout de termes de degrés supérieurs

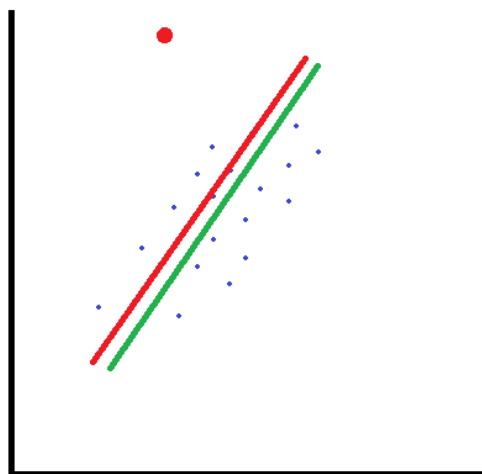


Non corrélation des résidus

- Les modèles supposent que les résidus ne sont pas corrélés: $Cov(e_i, e_j) = 0, i \neq j$
- Lorsque ce n'est pas le cas, il y a auto-corrélation, ou dépendance entre les observations
- Supposée pour les principales analyses statistiques, sauf les analyses pour données pairées (Mc Nemar, T Pairé, Wilcoxon rangs signés) et les modèles mixtes ou à effets aléatoires
- Vérification graphique ou par test



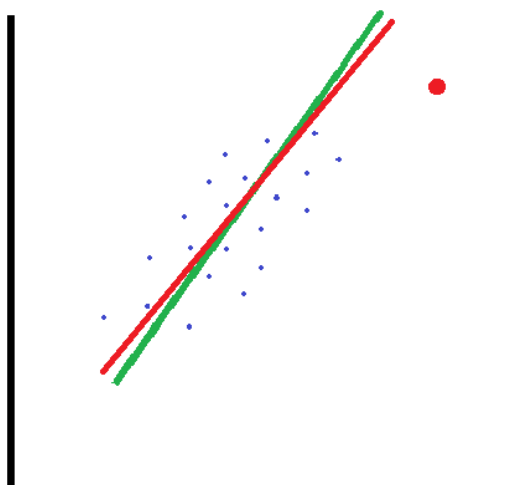
Valeurs influentes



Valeur aberrante

Observations ayant une combinaison de valeurs (VI, VD) très différente du reste des observations

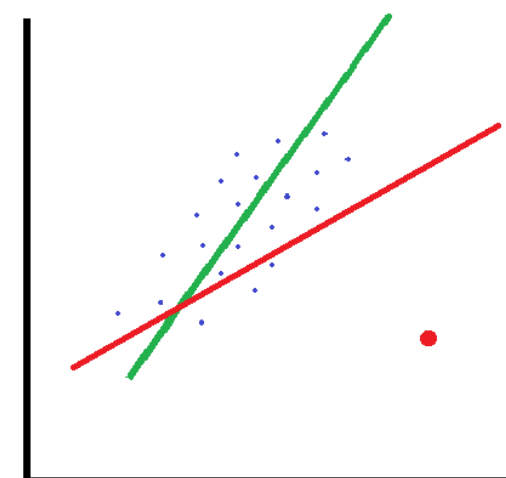
-> impact léger



Levier

Observation ayant une valeur de VI loin de la moyenne des VI

-> impact léger



Valeur influente

Observation avec levier et valeur VD différente des autres avec même VI

-> impact sur paramètres estimés et prédiction



Question 10

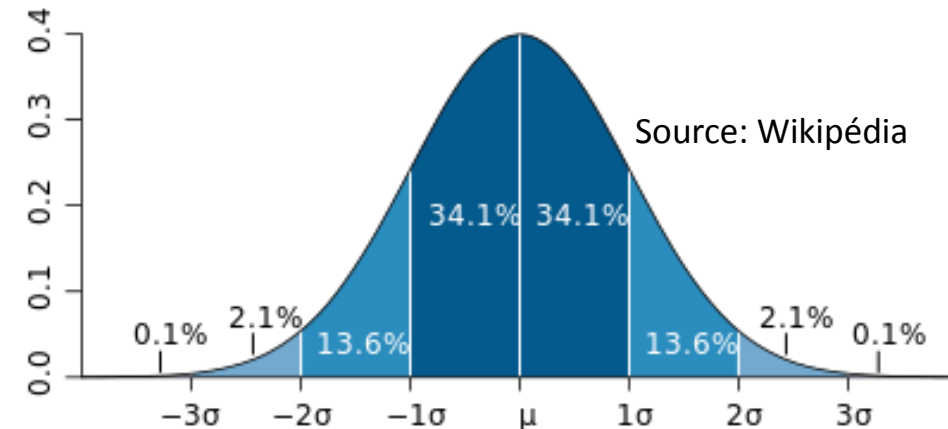
Quel test doit-on effectuer pour justifier l'exclusion de points singuliers? Dans quel contexte est-il acceptable d'exclure une valeur « aberrante »?

Par exemple, certains réplicats donnant une valeur dépassant d'au moins 2 fois l'écart type par rapport aux valeurs restantes. On voudrait pouvoir justifier leur exclusion afin de diminuer les écarts et pouvoir obtenir une significativité.



Réponse 10

- Attention, "pour obtenir une significativité"
- Distribution seule de la variable, pas suffisamment d'informations
- Groupes sous-jacents?
- Population cible?
- Postulats validés?
- Distribution normale, 95% entre +/-2 écarts-types





Valeurs influentes

Si l'analyse des résidus montre qu'il y a une valeur influente:

- Corriger la valeur s'il s'agit d'erreur de saisie
- Supprimer l'observation si elle ne fait pas partie de la population cible
- Analyse de sensibilité (avec et sans l'observation) si elle fait partie de la population cible



Questions 11

Quelles statistiques sont à utiliser avec les cultures cellulaires et un faible d'échantillon par groupe (triplicata)? Comment répartir les échantillons par plaque?

Quel test statistique utiliser pour comparer des traitements si on teste 2 molécules différentes à 3 doses en combinaison les unes avec les autres? Comment préparer l'expérience sur les animaux et quels tests sont à considérer pour analyser les effets de plusieurs traitements en prenant en compte le sexe d'animaux et le background génétique différent (wild type, KO).



Blocage

Utilisé lorsque qu'il pourrait y avoir de la variabilité quant à un facteur.

But: Corriger un facteur qui est hétérogène en le séparant en morceaux (blocs) afin d'avoir des unités homogènes à l'intérieur de chacun.

- Restriction à la randomisation
- Appliquer l'ensemble des traitements dans chacun des blocs homogènes, au lieu de les répartir aléatoirement dans l'ensemble du dispositif.
- Les traitements sont donc soumis à différentes conditions de façon équivalente.

Exemple:

Besoin de plusieurs plaques pour une expérience

Plaque 1: Contrôle, Plaque 2: Mutant... Différence entre Ctl et Mutant due à quoi?

Conditions pourraient varier d'une plaque à l'autre

S'assurer d'avoir tous les croisements possibles dans chaque plaque pour contrôler ce problème



Structure d'expérience

Structure des traitements

- Facteurs croisés ou emboîtés
- Effets fixes ou aléatoires
- Mesures uniques ou répétées

Structure de randomisation

PCA, PBAC, split plot, strip bloc...

Références:

- Montgomery, D.C. (2013). Design and Analysis of Experiments, 8th edition. Wiley. 752 pages.
- <http://www.tfrec.wsu.edu/ANOVA/index.html>



Types d'effets

Effet fixe

- Les modalités utilisées sont les seules d'intérêt pour lesquelles on désire obtenir une conclusion.
- Ces modalités ne pourraient pas être remplacées par d'autres sans changer l'objectif de recherche.

Effet aléatoire

- Les modalités utilisées pour l'effet aléatoire sont un échantillon de toutes celles disponibles.
- On ne désire pas comparer les modalités entre elles.
- On veut que les conclusions soient généralisables à leur population complète.
- Les modalités utilisées pourraient donc très bien être remplacé par d'autres.



Modèles linéaires mixtes

- Famille de modèles statistiques
- Estimation d'effets fixes et aléatoires
- Pour estimer l'ANOVA à mesures répétées, avec facteurs intra seulement ou intra+inter, l'ANOVA à effets aléatoires, les devis tels que split plot, PBAC, etc.
- Interaction entre tous les effets
- Permet de tenir compte de la dépendance entre les observations (pseudos répétitions) ou de contrôler pour des effets aléatoires (blocs)
- Très flexible

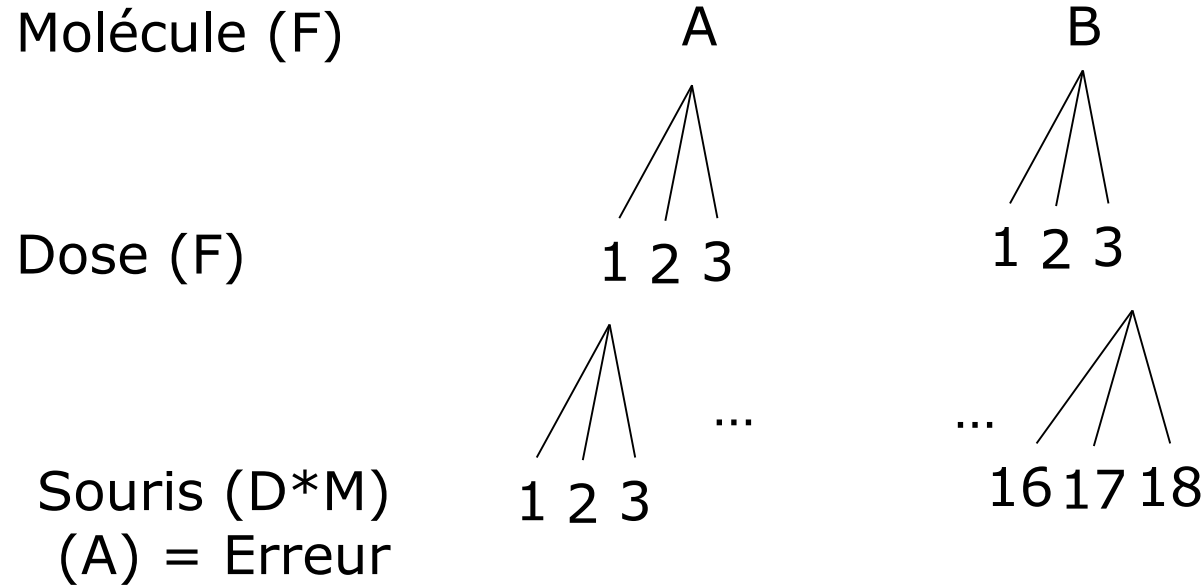


Réponse 11

- Selon le design expérimental, analyse statistique différente
- Arbre du design aide grandement
 - Ordre dans lequel les choses sont faites
 - Facteurs emboîtés / Croisés
 - Facteurs fixes / aléatoires
- Modèles linéaires mixtes suppose VD continue
- Postulats à valider
- Ajout du sexe et background génétique
 - Facteurs fixes ou aléatoires?
 - Où les ajouter dans l'arbre?



Design 1



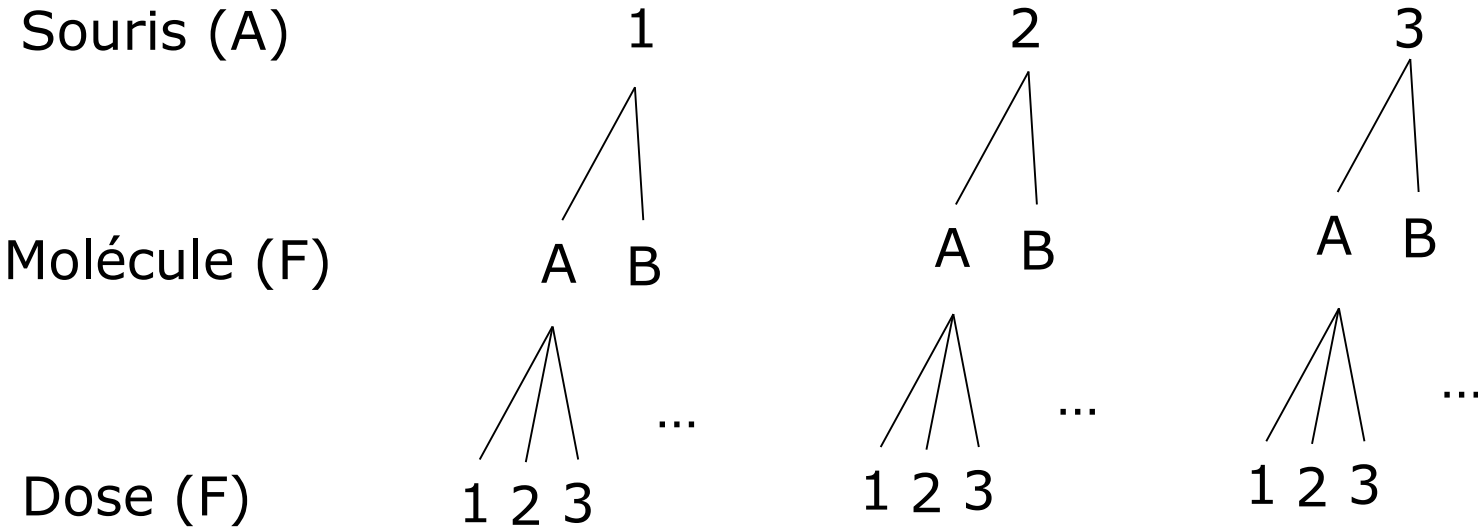
Plan factoriel à 2 facteurs croisés

Unité expérimentale = unité échantillonnale = souris

-> ANOVA à 2 facteurs (inter / fixes)



Design 2



Plan à blocs aléatoires complets (PBAC)

Unité expérimentale = unité échantillonnale = échantillon de souris

-> ANOVA à mesures répétées à 2 facteurs (intra / fixes)

-> Modèle linéaire mixte



Design 3

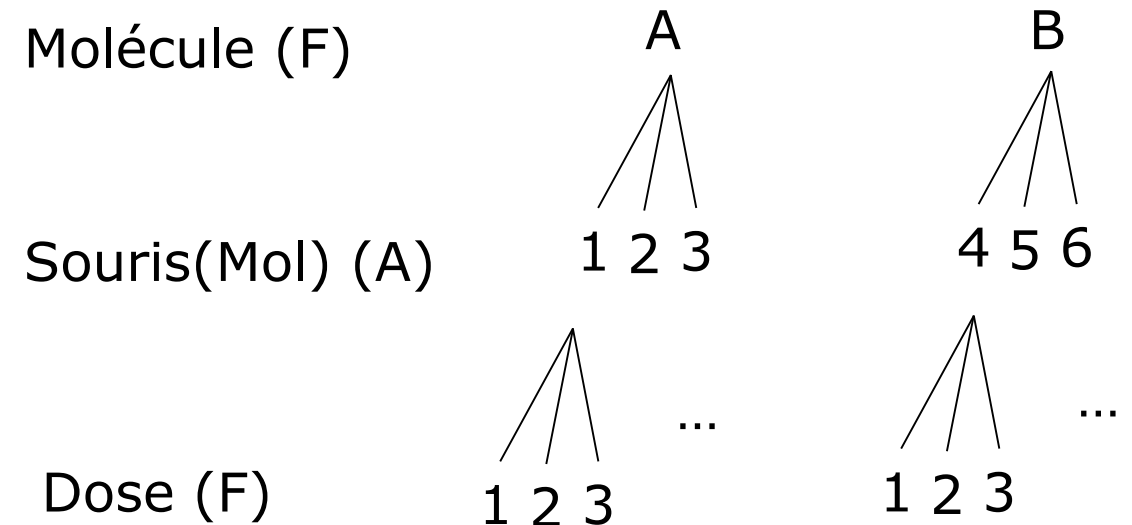
Split plot dans un plan aléatoire complet (PCA)

Deux unités expérimentales différentes:

- Molécule -> souris (parcelle principale)
- Dose -> échantillon de la souris (sous-parcelle)

-> ANOVA à mesures répétées à 2 facteurs (1 inter + 1 intra, fixes)

-> Modèle linéaire mixte





Autres points importants



Autres analyses

Pour tester l'association entre des variables qualitatives

Tests	Variable dépendante	Variable(s) indépendante(s)	Type de données	Type d'analyse
Khi-Deux	Nominale	Nominale	Non corrélées	Bivariée
Fisher Exact	Dichotomique	Dichotomique	Non corrélées	Bivariée
Mc Nemar	Dichotomique	Dichotomique	Corrélées	Bivariée
Régression logistique	Dichotomique	Qualitative(s) + Quantitative(s)	Non corrélées	Bivariée ou Multivariée

Pour quantifier l'association entre deux variables dichotomiques

- Rapport de cotes
- Rapport de risques



Autres analyses (suite)

Pour tester l'association entre des variables continues

Tests	Variable dépendante	Variable (s) indépendante(s)	Type de données	Type d'analyse
Corrélation	Continue	Continue	Non corrélées	Bivariée
Régression linéaire	Continue	Continue(s)	Non corrélées	Bivariée ou multivariée

Pour étudier le risque de survenue d'un événement

Tests	Variables dépendantes	Variable indépendante	Type de données	Type d'analyse
Kaplan-Meier	Continue (temps) + Dichotomique	Aucune ou Qualitative	Non corrélées	Univariée ou Bivariée



Interprétation des résultats

- Conclusion fiable seulement si les postulats des modèles sont respectés
- Une valeur $p > 0,05$ n'implique pas qu'il n'y a pas d'effet significatif -> ne rejette pas l'hypothèse nulle, examiner l'intervalle de confiance, vérification de la puissance
- La valeur p devrait être interprétée comme un continuum plutôt qu'une valeur dichotomique (présenter sa véritable valeur numérique, pas juste sig, ns, $p < 0.05$, $p > 0.05$...)
- La p value doit être accompagnée:
 - Mesures d'association (taille d'effet, rapport de risques, rapport de cotes, coefficient de régression ou corrélation, etc.)
 - Intervalles de confiance pour connaître la précision et l'incertitude des estimateurs ponctuels
- Tenir compte de la significativité clinique



Atelier 09 – Validation de modèles paramétriques

– Suite de l’atelier d’aujourd’hui

- Date: 8 juin 14h-16h
- Lieu: CHUL, local H00-329
- Contenu
 - Résidus
 - Postulats fréquents
 - Validation ANOVA, régression linéaire & régression logistique

- Inscription sur:

<http://rce.crchudequebec.ulaval.ca/>